

RESEARCH ARTICLE

Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis

Brenda K. Todd¹  | Rico A. Fischer²  | Steven Di Costa³ |
Amanda Roestorf¹ | Kate Harbour⁴ | Paul Hardiman⁵ | John A. Barry⁵ 

¹ Department of Psychology, City, University of London, London, UK

² Glasgow Caledonian University, Glasgow, UK

³ Institute of Cognitive Neuroscience, University College London, London, UK

⁴ City, University of London, London, UK

⁵ Institute of Women's Health, University College London, London, UK

Correspondence

Brenda K. Todd, Department of Psychology, City, University of London, London EC1V OHB, UK.

Email: brenda.todd.1@city.ac.uk

Abstract

From an early age, most children choose to play with toys typed to their own gender. In order to identify variables that predict toy preference, we conducted a meta-analysis of observational studies of the free selection of toys by boys and girls aged between 1 and 8 years. From an initial pool of 1788 papers, 16 studies (787 boys and 813 girls) met our inclusion criteria. We found that boys played with male-typed toys more than girls did (Cohen's $d = 1.03$, $p < .0001$) and girls played with female-typed toys more than boys did (Cohen's $d = -0.91$, $p < .0001$). Meta-regression showed no significant effect of presence of an adult, study context, geographical location of the study, publication date, child's age, or the inclusion of gender-neutral toys. However, further analysis of data for boys and girls separately revealed that older boys played more with male-typed toys relative to female-typed toys than did younger boys ($\beta = .68$, $p < .0001$). Additionally, an effect of the length of time since study publication was found: girls played more with female-typed toys in earlier studies than in later studies ($\beta = .70$, $p < .0001$), whereas boys played more with male-typed toys ($\beta = .46$, $p < .05$) in earlier studies than in more recent studies. Boys also played with male-typed toys less when observed in the home than in a laboratory ($\beta = -.46$, $p < .05$). Findings are discussed in terms of possible contributions of environmental influences and age-related changes in boys' and girls' toy preferences.

Highlights:

- Gender differences in toy choice exist and appear to be the product of both innate and social forces.
- Despite methodological variation in the choice and number of toys offered, context of testing, and age of child, the consistency in finding sex differences in children's preferences for toys typed

to their own gender indicates the strength of this phenomenon and the likelihood that has a biological origin.

- The time playing with male-typed toys increased as boys got older, but the same pattern was not found in girls; this indicates that stereotypical social effects may persist longer for boys or that there is a stronger biological predisposition for certain play styles in boys.

KEYWORDS

gender differences, play, toy preference

1 | INTRODUCTION

Decades of study, beginning with Benjamin's (1932) paper, find evidence of sex differences in children's toy preference; data have been collected in a range of social contexts, using visual preference and experimental and observational methodologies. Sex differences in children's object preferences may originate in biological predispositions, which are subsequently influenced by social processes. Alternatively, they may be solely or primarily attributed to social factors. In either case, the impact of socialization is likely to be modified as cognition develops and boys and girls become aware of their gender group affiliation and associated norms.

Despite a wealth of pertinent data, there is still uncertainty about the degree to which given variables contribute to outcomes and vary with child age. This topic is of theoretical and everyday interest as parents, educators, marketers, and the media question how children's gender-typed interests are best supported or diverted.

Hyde (2005) recommends meta-analyses to examine the influence of context on behavioural sex differences and in identifying age trends in their magnitude, yet her review of meta-analyses pertaining to sex differences does not cover sex differences in children's play. We therefore present a meta-analysis of systematic observations of children's free selection of toys spanning several decades. Our aim is to estimate the effect of the social and cultural contexts of testing, child age, and methodologies where boys and girls are offered gender-neutral as well as gender-typed toys. Whereas many meta-analyses focus on small sex differences, we aim to investigate whether sex differences in toy choice are as large as they are perceived to be and as reported to be in some individual studies.

1.1 | Biological predispositions

Aptitudes that may impact on boys' toy choice include their typical advantage over girls in gross motor skills (Touwen, 1976) and propulsive movement (Benenson, Tennyson, & Wrangham, 2011), higher activity levels (Campbell & Eaton, 1999), and lower impulse control (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006). Toys affording construction and mechanical movement may appeal more to boys because of their early advantage in mental rotation of figures (Moore & Johnson, 2008; Quinn & Liben, 2008, 2014) and event mapping (Schweinle & Wilcox, 2004; Wilcox, 2003). In contrast, girl's greater attraction to social stimuli may account for their preference for toys that afford nurturance; girls engage more than boys in mutual gaze (Lavelli & Fogel, 2002; Leeb & Rejskind, 2004) and have a small advantage in processing facial expressions (McClure, 2000). Girls' typical advantage in fine motor finger control (Nagy, Kompagne, Orvos, & Pal, 2007; Touwen, 1976) may also drive toy selection.

The interests and aptitudes that pertain to differential attraction to objects may relate to hormonal characteristics. A large body of evidence from research with mammals shows strong effects of exposure to prenatal and perinatal sex hormones on sex-related behaviours, including juvenile play (Meyer-Bahlburg et al., 2004). Both prenatally and

during the first 6 months of life, boys are typically exposed to higher levels of androgens than girls, resulting in masculinization of the neural systems and of behaviour (Collaer & Hines, 1995; Hines, 2004). There is some evidence that levels of androgen exposure affect object preference in very young infants; for example, in a study of the visual preferences of 3–8-month-old infants, boys made more fixations to a truck than a doll ($d = 0.78$), whereas girls had more interest in the doll than the truck ($d > 1.0$) (Alexander, Wilcox, & Woods, 2009). Stronger evidence of hormonal influence on toy preference comes from clinical populations (e.g., Cohen-Bendahan, van de Beek, & Berenbaum, 2005; Hines, Brook, & Conway, 2004). A review of studies of the effect of testosterone concentrations on human development indicates the contribution of this hormone to the masculinization of behaviour (Alexander, 2014). Girls with congenital adrenal hyperplasia who are exposed to higher levels of androgens than other girls show correspondingly greater interest in male-typed over female-typed toys (Hines, 2004; Meyer-Bahlburg et al., 2004) and girls more affected by congenital adrenal hyperplasia showed greater interest in male-typed toys than those less affected by the condition (Van de Beek, van Goozen, Buitelaar, & Cohen-Kettenis, 2009).

Further indications of relevant biological predispositions arise from the findings of sex differences in object preference among nonhuman primates: female rhesus monkeys make more contact with a doll and a cooking pot than males do, whereas males spend longer than females manipulating a car and a ball (Alexander & Hines, 2002), and male rhesus monkeys prefer wheeled over nonwheeled plush toys (Hassett, Siebert, & Wallen, 2008).

In general, biological explanations predict continuity in boys' and girls' object preferences over late infancy and childhood and across contexts of testing.

1.2 | The effects of socialization and study setting

Various environmental processes involving the attitudes and behaviour of parents and other children may modify children's toy preferences and is likely to vary by child age and cultural norms.

Although parents' reported opinions on gender stereotypes do not correlate strongly with their children's behaviour (Fagot, 1974; Smith & Daglish, 1977), parent behaviour may; a meta-analysis indicates that parents encourage gender-typed activities, though decreasingly so as children grow older (Lyttton & Romney, 1991). A social context where a parent is present may therefore incline a younger child to exhibit stereotypical preferences. However, some recent studies show that infants engage more in gender-typed play when alone than with a parent (Zosuls et al., 2009), as do typically developing girls, but not boys (Pasterski et al., 2005). The effect of socialization may depend on parent and child sex and vary as parents redefine their roles in interaction with the knowledge and perceived needs of younger and older children.

When the play partner is another child, gender-typed play is typically increased; preschool children spend less time playing with "other-gender-typed" toys with a peer than in solitary play (Serbin, Connor, Burchardt, & Citron, 1979), and children's behaviour is more highly stereotyped in group than dyadic situations (Banerjee & Lintern, 2000; Fabes, Martin, & Hanish, 2003). Children engage more with same-sex than other-sex peers (Hines & Kaufman, 1994; Martin & Fabes, 2001), and in such contexts, their toy preferences may reflect their motivation to conform to sex role stereotypes as much as individual preference.

Typically, parents' differential socialization of boys and girls and children's own attitudes are likely to reflect their cultural context. Lyttton and Romney's (1991) meta-analysis of parental socialization showed differences between North American studies and those from other Western countries. Few individual studies of children's toy preference include regional comparisons. Although Turner and Gervai (1995) found substantial replication of parental effects on toy preference between the United Kingdom and Hungary, the finding that both girls and boys in Sweden preferred female-typed toys less as age increased (Servin, Bohlin, & Berlin, 1999) contrasts with similar but earlier studies, conducted in the United States, which found increases in girls' preference for female-typed toys with age (Blakemore, LaRue, & Olejnik, 1979; O'Brien & Huston, 1985). Servin et al. (1999, p. 40) speculated that the difference they observed may be explained by Sweden's "equal-roles family model."

The study of children's toy preferences has spanned several decades, allowing for a comparison of boys' and girls' behaviour in different historical time periods. Historical changes in gender roles are likely to impact on socialization;

Karraker, Vogel, and Lake's (1995) replication of a study by Rubin, Provenzano, and Luria (1974) found a reduction in U.S. parents' stereotypical perception of newborns; similarly, Signorella, Bigler, and Liben (1993) found smaller age and sex effects in children's own nonstereotyped responses in earlier relative to later studies (conducted between 1963 and 1991). However, Zosuls et al. (2009) suggest that many Western middle-class parents currently promote gender-neutral play. We propose, therefore, the effect of time period can be used as a proxy measure for the effect of culture on children's gender-typed object preferences.

The immediate settings in which children's toy preferences are tested vary in the formality and social context of testing, factors that might affect behaviour. For example, laboratory and home contexts typically involve interaction with, or the presence of, a parent, whereas in school or nursery settings, the focal child may be in the presence of their peers.

Comparing findings from studies conducted in different social settings (e.g., presence/absence of parent or peers) as well as geographical and historical contexts, while controlling for other factors, may inform us on specific environmental effects on children's sex-typed choices. Extrapolating from the studies described above, we might therefore expect to see the magnitude of sex differences in toy preference differ by (a) social context of testing; (b) geographical region; (c) date since publication; and (d) child age. One specific aspect of context that has not received research attention is whether the immediate environmental setting of the study (home, nursery, laboratory, etc.) influences boys' and girls' toy choice.

1.3 | The development of gender-typed behaviour

Throughout this review, we refer to the potential impact of child age on the degree to which gender-typed preferences for toys are demonstrated with the assumption that a child's conception of gender and associated stereotypes will expand with age.

Conceptualization of gender is comprised of perceptual, cognitive, and social components, acquired incrementally. By 6 months of age, infants distinguish between male and female faces (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002) and voices (Miller, 1983). Explicit labelling of others by gender occurs between the ages of 18–24 months (for a review, see Martin & Ruble, 2010), and it is likely that infants begin to understand that they too have a gendered identity during this period. As the concept of gender develops, selective modelling (Slaby & Frey, 1975) of parent and peer behaviour is expected to increase and the media become more influential (Diekman & Murnen, 2004; Leaper, Breed, Hoffman, & Perlman, 2002). Higher levels of gender category knowledge predict increased awareness of stereotypical preferences.

In contrast to the development of stereotyped cognition, Martin and Ruble (2004) argue that gender-typed behaviour shows more mixed developmental patterns. A meta-analysis of studies of developmental change in gender schemata shows that, although knowledge of gender stereotypes increases with age, behaviour in accordance with the stereotypes does not necessarily follow (Signorella et al., 1993), and this aspect requires further exploration.

Studies that include infants and/or preschool children in more than one age group generally report increases in gender-typed toy preference with age (e.g., O'Brien, Huston, & Risley, 1983; Todd, Barry, & Thommessen, 2016; Zosuls et al., 2009; Zosuls, Ruble, & Tamis-LeMonda, 2014), but the distinct developmental patterns of boys and girls vary between studies (Servin et al., 1999). Drivers of developmental change are multifaceted, and a more systematic comparison of boys' and girls' stereotypical preferences at different ages and across different social settings and cultural contexts may add to our understanding.

1.4 | Methodological aspects of previous research and rationale for choice of studies in the present meta-analysis

Toy preferences have been measured in a variety of ways. Parental and child self-report, sometimes retrospective, is valuable when working with large samples but may be prone to bias in recall (Yarrow, Campbell, & Burton, 1970) or demand characteristics (see Wilansky-Traynor & Lobel, 2008, for a discussion). Visual preference techniques can test infants whose motor development is insufficient to demonstrate preference in other ways. However, it may be the

child's ability to act on an object, as well as its visual properties, that is appealing. This meta-analysis covers only observational studies testing children's free preferences in various controlled settings because we believe that these best indicate a child's free preference in a play situation. Meta-regression is used to control for any influence of child age and study context on toy preference. Where extraneous variables could not be controlled in any meaningful way, studies were excluded; for example, we exclude studies measuring play with peers because competition for toys may limit choice: studies involving play with the child's own toys, on the basis that available toys differ between participants and are already familiar to them. Included studies use a wide range of gender-typed toy stimuli, and many also include toys defined as gender-neutral. This variation in methodology is also controlled as it may impact the effect size of found sex differences.

In summary, studies reporting observations of children in free play were considered the best indicator of toy preference because (a) they typically offer a range of gender-typed toys and (b) they allow manipulation and movement of the stimuli and have the potential for symbolic play.

1.5 | The present review

This meta-analysis was designed to test the following hypotheses:

1. Gender-typed preferences will be observed in children's free play with toys.
2. The setting in which the child plays, for example, the presence of a parent or location in a laboratory rather than home, will affect the magnitude of gender-typed behaviour.
3. Effect sizes for toy preference will be smaller in more egalitarian compared with less egalitarian societies.
4. Effect sizes for toy preference will be larger in historically earlier studies rather than later studies.
5. Sex differences in toy preference will increase with age.
6. Effect sizes for toy preference will be smaller when the children are presented with gender-neutral toys at the same time as they are presented with gender-typed toys.

2 | METHOD

2.1 | Selection of studies

The selection of studies followed the guidelines issued by Stroup et al. (2000). Qualified librarians advised on the search strategy and assisted where papers were difficult to access. The study protocol is available from the authors.

2.2 | Inclusion and exclusion criteria

Five inclusion criteria were applied: first, studies included typically developing children; second, duration of playing with a toy was reported as mean and standard deviation or standard error in seconds, or this information could be extracted indirectly from the paper, for example, from a histogram with error bars; third, the study provided observational data or provided baseline data prior to experimental manipulation; fourth, the study reported other relevant data, for example, the number of children in each group; fifth, stimuli toys were equally available to all participants and not shared with other children at the time of testing.

Papers not published in English were translated and reviewed.

Published studies were assessed from the earliest possible dates for each database up to 24 September 2016. The search terms used, in topic, were ("toy choice*" OR "toy preference*" OR "toy play") OR ("toy" AND "play behav*") OR ("gender typ* play") OR ("sex-typ* play") OR ("gender stereotyp* toy") OR ("gender stereotyp* play") OR ("sex-typ*

toy") OR ("sex stereotyp* toy") OR ("gender typ* activit*") OR ("sex typ* activit*"). As Figure 1 shows, this yielded 520 hits from the Web of Knowledge or Web of Science (from 1900), 205 from Embase Classic + Embase (from 1947), 11 from Maternity and Infant Care (from 1971), 200 from Ovid MEDLINE(R) (from 1946), and 852 from PsycINFO (from 1806). A hand search of the resulting papers produced no further relevant papers. Authors were contacted where additional information was needed. For example, the Lamminmäki et al. (2012) study did not include the mean and SD seconds of play, so the authors of that paper were contacted and provided this information.

2.3 | Data extraction

Data were collected and coded based on relevance to child development research and toy choice paradigm methodology. Five variables were identified as having a possible impact on the child's toy choice and were used as covariates in the meta-regression. The following data were extracted from each study:

- (i) The presence of another person during the play task. This was categorized such that higher scores indicate more potential influence of the presence of another person on the child: 1 = child completely alone; 2 = child engaged in sole play with minimal parental or researcher interaction; 3 = combination of 50% sole play with inactive parent and 50% parent present and actively involved in play with child.

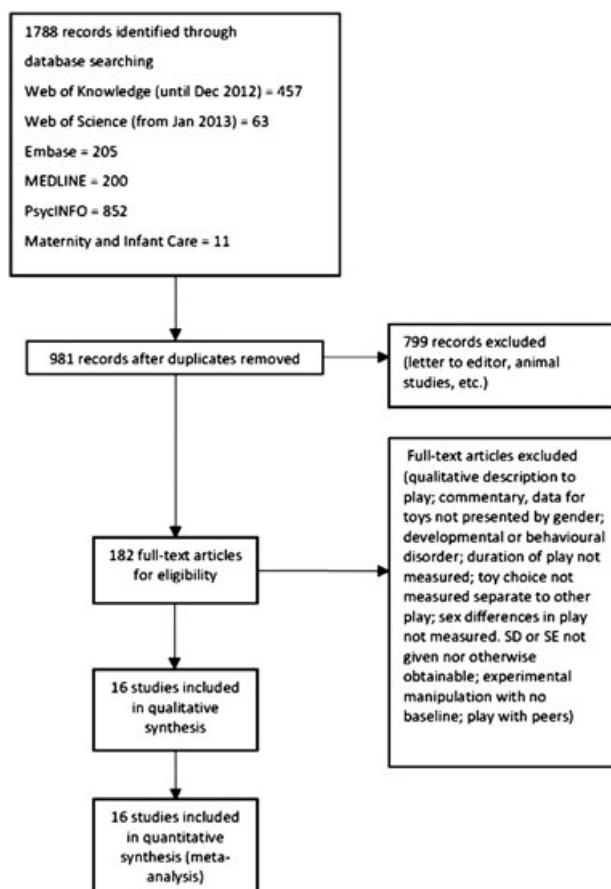


FIGURE 1 Search process flow chart from initial study pool to final selection of studies

TABLE 1A Information extracted from the included papers with reference to male-typed toys

Study	Boys			Girls			Neutral toys SD	Neutral toys Year	Publication years	Parent present	Setting	Country
	Age (months)	N	Mean play	SD	N	Mean play						
Alexander and Saenz (2012)	19.5	47	438.72	166.08	47	248.64	161.28	1	2012	3	3	2
Arthur (2014)	19	22	148	105	18	54	53	0	2014	2	2	2
Berenbaum and Hines (1992)	65.79	18	325.25	270.21	18	135.13	201.72	1	1992	23	2	2
Berenbaum and Snyder (1995)	93.1	19	484.21	141.28	19	202.48	168.14	1	1995	20	2	1
Doering, Zucker, Bradley, and Machtyre (1989)	87	15	260.1	67.7	15	98.4	101.3	1	1989	26	1	2
Idle, Wood, and Desmarais (1993)	46	10	393.2	136.2	10	271.2	139.7	1	1993	22	3	3
Lamminmäki et al. (2012)	14	21	46.2	59.67	21	33.54	47.16	1	2012	3	2	2
Li and Wong (2016)	77	38	214.7	132.26	38	43.38	85.79	1	2016	0	1	1
Pasterski et al. (2005)	81	12	244.8	118.8	12	7.2	14.4	1	2005	10	3	2
Pasterski et al. (2005)	81	25	144	144	25	43.2	104.4	1	2005	10	3	2
Pasterski et al. (2005)	81	24	252	100.8	24	46.8	93.6	1	2005	10	1	2
Schau, Kahn, Diebold, and Cherry (1980)	48	26	357.04	209.02	26	131.48	107.9	1	1980	35	2	2
Servin et al. (1999)	12	19	182.29	103.93	19	114.48	75.02	1	1999	16	3	2
Servin et al. (1999)	36	13	279.88	67.09	13	139.75	77.45	1	1999	16	3	2
Servin et al. (1999)	60	14	244.07	123.3	14	150.42	104.55	1	1999	16	3	2
Todd et al. (2016)	13	20	31.65	29.15	20	5.75	11.55	0	2014	0	2	3
Todd et al. (2016)	21.5	18	46.95	29.7	18	12.3	17.65	0	2014	0	2	3
Todd et al. (2016)	28	16	75	28.6	16	13.6	17.6	0	2014	0	2	3
Van de Beek et al. (2009)	13	63	169.98	118.23	63	112.47	94.54	1	2009	6	2	2
Wilansky-Traynor and Lobel (2008)	65.83	26	158.33	80.79	26	80.45	66.38	0	2008	7	2	3
Wilansky-Traynor and Lobel (2008)	65.83	31	130.89	89.36	31	82.01	75.83	0	2008	7	1	3
Zosuls et al. (2009)	18	36	72	90	36	30	36	1	2009	6	2	1
Zosuls et al. (2009)	22	36	78	114	36	42	42	1	2009	6	2	1
Zosuls et al. (2009)	18	36	54	84	36	42	54	1	2009	6	3	1
Zosuls et al. (2009)	22	36	66	66	36	36	84	1	2009	6	3	1
Zosuls et al. (2014)	24	93	64.8	56.7	99	32.4	35.1	1	2014	2	2	2
Zosuls et al. (2014)	36	63	67.5	48.6	77	43.2	43.2	1	2014	2	2	2

Note. N = sample size; Mean play = mean seconds played with toy; SD = standard deviation; Neutral toys (1 = neutral toys included; 2 = neutral toys not included); Publication years = years since publication of the study; Parent present (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place. The raw data for this table can be accessed at <http://dx.doi.org/10.6084/m9.figshare.5047660>

TABLE 1B Information extracted from the included papers with reference to female-typed toys

Study	Boys				Girls				Neutral toys	Year	Publication years	Parent present	Setting	Country
	Age (months)	N	Mean play	SD	N	Mean play	SD							
Alexander and Saenz (2012)	19.5	47	377.28	170.88	37	538.56	214.08	1		2012	4	3	2	0.26
Arthur (2014)	19	22	126	67	18	244	54	0		2014	2	2	2	0.28
Berenbaum and Hines (1992)	65.79	18	33.5	69.55	15	247.93	246.5	1		1992	24	2	2	0.25
Berenbaum and Snyder (1995)	93.1	19	26.42	91.95	13	265.42	240.71	1		1995	21	2	1	0.25
Doering et al. (1989)	87	15	7.2	21.5	15	145.2	109.6	1		1989	27	1	2	0.14
Idle et al. (1993)	46	10	136.6	92.4	10	140.5	126.8	1		1993	23	3	3	0.14
Lamminmäki et al. (2012)	14	21	80.65	68.63	26	80.09	79.6	1		2012	4	2	2	0.25
Li and Wong (2016)	77	38	26.96	62.5	38	183.28	117.5	1		2016	0	1	3	
Pasterski et al. (2005)	81	12	21.6	36	12	115.2	169.2	1		2005	11	3	2	0.23
Pasterski et al. (2005)	81	25	43.2	93.6	27	147.6	169.2	1		2005	11	3	2	0.23
Pasterski et al. (2005)	81	24	21.6	68.4	26	219.6	136.8	1		2005	11	1	2	0.23
Schau et al. (1980)	48	26	92.08	102.25	26	166.92	159.62	1		1980	36	2	2	0.29
Servin et al. (1999)	12	19	220.82	103.27	19	275.08	98.2	1		1999	17	3	2	0.07
Servin et al. (1999)	36	13	73.19	56.9	18	215.83	156.92	1		1999	17	3	2	0.07
Servin et al. (1999)	60	14	23.75	44.08	21	139.1	110.38	1		1999	17	3	2	0.07
Todd et al. (2016)	21.5	18	7.35	13.45	11	42.05	38.35	0		2014	2	2	3	0.19
Todd et al. (2016)	28	16	6.1	7.4	16	57.35	28.95	0		2014	2	2	3	0.19
Todd et al. (2016)	13	20	14	17.65	20	58.9	31.85	0		2014	2	2	3	0.19
Van de Beek et al. (2009)	13	63	151.97	117.28	63	189.82	112.96	1		2009	7	2	2	0.05
Wilansky-Traynor and Lobel (2008)	65.83	26	63.43	72.29	30	148.3	77.38	0		2008	8	2	3	0.14
Wilansky-Traynor and Lobel (2008)	65.83	31	61.52	74.03	29	125.03	79.32	0		2008	8	1	3	0.14
Zosuls et al. (2009)	18	36	78	120	46	90	108	1		2009	7	2	1	0.29
Zosuls et al. (2009)	22	36	54	60	46	150	162	1		2009	7	2	1	0.29
Zosuls et al. (2009)	18	36	96	102	46	108	116	1		2009	7	3	1	0.29
Zosuls et al. (2009)	22	36	78	84	46	120	136	1		2009	7	3	1	0.29
Zosuls et al. (2014)	24	93	29.7	40.5	99	62.1	64.8	1		2014	2	2	1	0.28
Zosuls et al. (2014)	36	63	32.4	43.2	77	75.6	72.9	1		2014	2	2	1	0.28

Note. N = sample size; Mean play = mean seconds played with toy; SD = standard deviation; Neutral toys (1 = neutral toys included; 2 = neutral toys not included); Publication years = years since publication of the study; Parent present (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place. The data for this table can be accessed at <http://dx.doi.org/10.6084/m9.figshare.5047663>

- (ii) The study location. Whether the study was set in the home or at another location was coded as follows: 1 = home; 2 = laboratory; 3 = nursery.
- (iii) The gender equality values of the region where a study was conducted. This was estimated using the gender inequality index (GII; Ferrant, 2010) in order to assess any effect of cultural gender stereotyping on children's toy preferences. GII scores represent gender inequality in reproductive health, parliamentary seats, higher educational achievement, and participation in the workforce. GII scores exist for 137 countries, from 1995 in most cases (United Nations Development Programme, n.d.). Lower scores indicate more gender equality. To give an idea of the range of scores, as of the latest date (2013), Sweden is third in the world table with a score of 0.05, and Yemen is at the bottom with 0.73. The GII is produced every few years and does not tend to vary much over time; for example, though the GII is falling in the United Kingdom over the five years measured (2000 = 0.23; 2005 = 0.22; 2010 = 0.21; 2012 = 0.21; 2013 = 0.19), this is a relatively slow decline. In the present meta-analysis, where there was no GII score in the year a study was conducted, the GII score in the year closest to the study was used.
- (iv) Years since publication of a study. This measure tests the hypothesis that the prevalence of gender stereotyping has changed over time with the expectation that older studies would find more stereotypical toy preference.
- (v) The age of the children. This information was extracted because of the relevance of children's developmental stage to preference for toys. For example, Zosuls et al. (2009) found that sex difference in toy preference was higher in older than in younger infants.
- (vi) The inclusion or not of toys defined as gender neutral. This variable tests for an effect of including gender-neutral toys on sex differences in toy preference and was coded as follows: 1 = included gender-neutral toys; 0 = no gender-neutral toys included.

The outcome variable was operationalized in three ways, reflected in Sections 3.2, 3.3, 3.5, and 3.6. First, in the meta-analyses and meta-regressions, the outcome was expressed as the standard mean difference between the duration with which boys and girls played with male- or female-typed toys. The effect size of this difference was estimated in units of Cohen's *d*. A Cohen's *d* of 0.2 is considered a small difference between groups, 0.5 a moderate difference, and 0.8 or more a large difference. Second, regressions were performed separately for boys and girls, with outcomes expressed as the percentage of time played with male-typed toys relative to female-typed toys. Third, the actual duration of play—rather than proportion or percentage of play—for each toy was assessed for boys and girls separately. Thus, the raw number of seconds of play with male- and female-typed toys, rather than the percentage of play with male toys relative to female toys, was assessed.

Tables 1a and 1b show information regarding (a) characteristics of the participants (mean age, sex); (b) the setting (presence of others; at child's home or nursery, etc.); (c) the country in which the study was conducted; (d) whether or not gender-neutral as well as male- and female-typed toys were included; (e) mean seconds played with toys; and (f) years since study publication. Table 2 shows additional data that were extracted in order to assess the methodological quality of each study, for example, whether a study had controlled for relevant variables.

Each article was assessed by authors S.D., A.R. or K.H., B.T., R.F., and J.B. Articles that fitted the main criteria (observing sex differences in children's choice of gender-typical toys) were accessed. Methodological quality was assessed by S.D., A.R., K.H., and R.F. and checked by J.B. based on the criteria of the Newcastle-Ottawa Quality Assessment Scale (NOS) for case-control studies (Wells et al., 2011). The Cochrane Nonrandomized Studies Methods Working Group considers the NOS an acceptable tool for assessment of nonrandomized studies (Reeves & Higgins, 2011). The criteria considered were (a) clear justification for gendered nature of toy, for example, based on research; (b) recruitment of consecutive participants; (c) whether boys and girls were comparable in terms of social background; (d) whether parents' views on gender were measured; (e) whether the toys were comparable (in size, shape, etc.) and the boys and girls ages were comparable; (f) clearly defined play behaviour; (g) clearly defined measurement of the outcome (time spent playing with toy); and (h) whether nonuptake or dropout rates were reported. Interrater

TABLE 2 Evaluation of the methodological quality of the 16 studies comparing toy preference in boys and girls in studies included in the meta-analysis

Study	Case definition adequate	Representativeness of cases	Selection of controls	Parental opinion	Comparability of both groups	Ascertainment of behaviour	Same ascertainment method for both groups	Non-response rate	NOS score
Schau et al. (1980)	*	*	*	*	*	X	*	X	6
Doering et al. (1989)	*	*	X	*	*	*	*	X	6
Berenbaum and Hines (1992)	X	*	*	*	*	*	*	X	6
Idle et al. (1993)	*	*	*	*	*	*	*	*	8
Berenbaum and Snyder (1995)	*	X	*	*	*	*	*	*	7
Servin et al. (1999)	*	*	X	*	*	*	*	*	7
Pasterski et al. (2005)	*	*	*	*	*	*	*	X	7
Wilansky-Traynor and Lobel (2008)	*	*	X	*	*	*	*	X	6
Zosuls et al. (2009)	*	*	X	*	*	*	*	*	7
Van de Beek et al. (2009)	*	*	X	**	*	*	*	*	8
Lamminmäki et al. (2012)	*	*	*	*	*	*	*	X	7
Alexander and Saenz (2012)	*	X	X	*	*	*	*	X	5
Zosuls et al. (2014)	*	X	X	*	**	*	*	*	7
Arthur (2014)	*	X	X	**	*	*	*	*	6
Todd et al. (2016)	*	*	X	*	*	*	*	*	7
Li and Wong (2016)	X	X	X	**	*	*	*	*	6

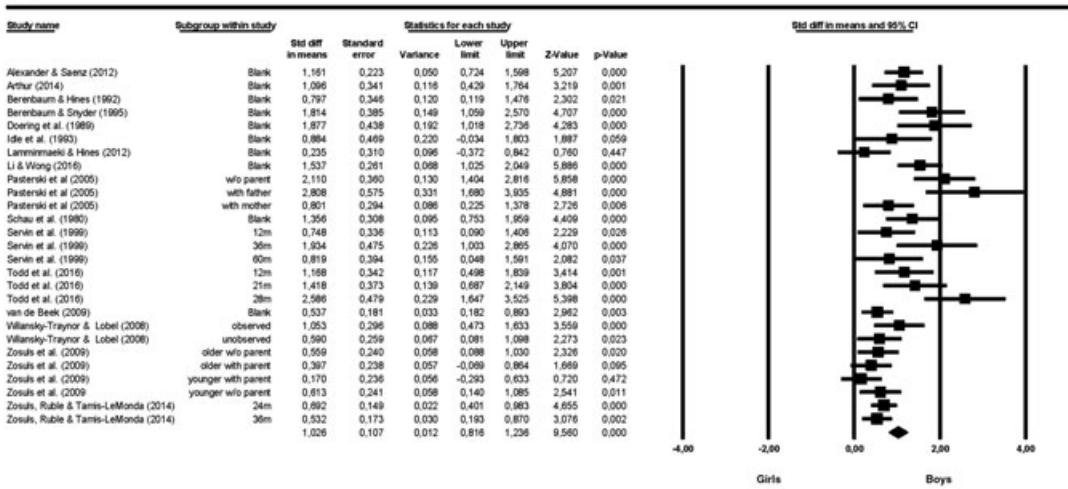
Note. Quality is assessed using Newcastle–Ottawa Quality Assessment Scale criteria adapted for this study. A star indicates that the study fulfilled this criterion; an X indicates that the study did not fulfil this criterion. Case definition adequate: clear justification for the gendered nature of a toy, for example, based on research. Representativeness of cases: recruitment of consecutive participants. Selection of controls: whether boys and girls were comparable in terms of social background. Parental opinion: whether parents' views on gender were measured. Comparability of both groups: the toys were comparable (in size, shape, etc.) and if the boys and girls were comparable in age. Ascertainment of behaviour: Play behaviour was clearly defined. Same ascertainment method for both groups: The measurement of the outcome (time spent playing with toy) was clearly defined. Nonresponse rate: whether either nonuptake or dropout rates reported.

agreement of NOS scoring was assessed using Cohen's kappa. A kappa of 0.41–0.60 is generally considered as demonstrating moderate agreement and 0.61 to 0.80 as substantial agreement (Landis & Koch, 1977). The NOS scores of S.D. and A.R. and R.F. and K.H. were in substantial agreement, as shown by a Cohen's kappa of 0.77 ($p = .000013$); the mean score for the 16 studies was 6.63 ($SD = 0.81$). Discrepancies were examined and discussed before a final score was agreed upon and assigned.

The time difference in boys' and girls' play with gender-typed toys in each age group in each study (shown in Figure 2a, b) was calculated as a common unit, Cohen's d . The combined effect size for all studies was calculated as the estimated treatment difference, Z , with a fixed- or random-effects model, as appropriate. Results were considered statistically significant where the probability value was below the .05 threshold. Statistical analyses were performed using Review Manager Version 5.1 (The Cochrane Collaboration, 2011), Stata Version 13.0 (StataCorp, Texas), and SPSS 22.

Heterogeneity was assessed using I^2 and chi-square statistics. I^2 values of 30% or above were considered likely to represent moderate heterogeneity, and chi-square p values $<.10$ were considered to represent significant heterogeneity; thus,

(a)



(b)

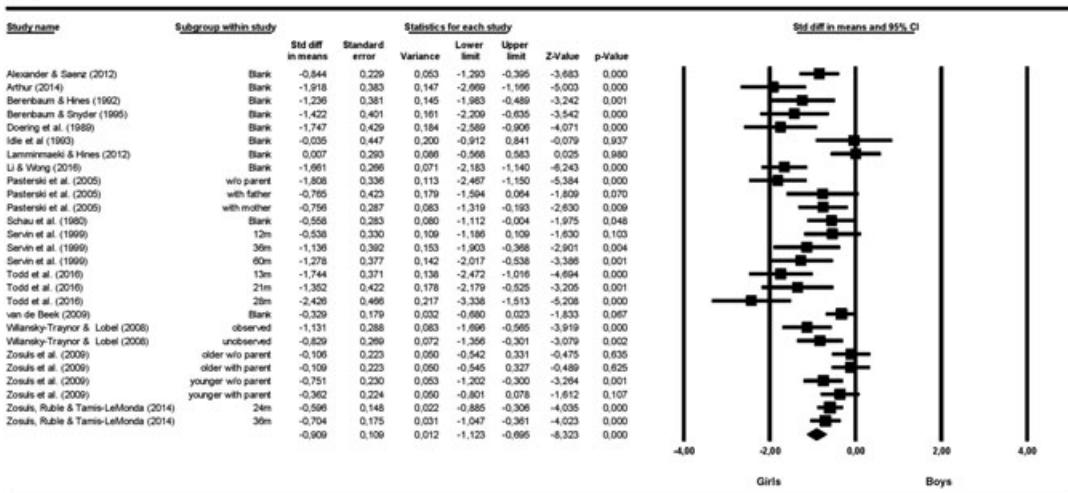


FIGURE 2 Forrest plot of meta-analysis of sex differences in children's play behaviour with (a) male-typed toys (b) female-typed toys

studies showing tolerable heterogeneity ($I^2 < 30\%$ and chi-square $p > .10$) were analysed using fixed-effects models, and studies beyond these limits ($I^2 > 30\%$ and chi-square $p < .10$) were analysed using random-effects models.

2.4 | Eligible articles

Figure 1 summarizes the search strategy used to identify appropriate studies. After duplicates were removed, the titles and abstracts of 961 records, from the 1788 initially retrieved, were assessed. A total of 182 full texts were further assessed, including reference sections, and 166 were excluded for various reasons (e.g., toy play data were not separated by child sex). One study was excluded (Berenbaum & Hines, 1992) because of overlap of participants with another (Berenbaum & Snyder, 1995); the more recent study was selected for inclusion because of the greater number of participants and the higher rating on the NOS scale.

Finally, 16 papers met all of the inclusion and exclusion criteria.

Some studies presented data in proportions of total time playing (e.g., Zosuls et al., 2009) or intervals of time (e.g., Todd et al., 2016), and these studies were included if the overall time in seconds was stated in the paper, from which an approximation of the mean time could then be calculated.

Some studies were excluded, for example, where there was an experimental manipulation with no baseline data given (e.g., Wolf, 1973) or where a child was assessed playing together with other children (e.g., Serbin, Connor, & Citron, 1981) or when children were tested at home with their own toys rather than with toys supplied by the experimenter (e.g., Fagot & Leinbach, 1989).

In one study (Pasterski et al., 2005), 84 (71%) of the participants were assessed in the United Kingdom, and 34 (29%) were assessed in the United States. In this case, the GII score for each country (0.22 for the United Kingdom and 0.29 for the United States) was weighted according to the relative number of participants in each country, giving a combined GII score for the study of 0.23.

The 16 studies that met the inclusion criteria for the meta-analysis (787 boys and 813 girls, in 27 groups) are listed in Tables 1a and 1b (raw data publicly available at <http://dx.doi.org/10.6084/m9.figshare.5047660> and <http://dx.doi.org/10.6084/m9.figshare.5047663>). The children had a mean (standard deviation) age of 43.25 (27.62) months old (minimum age group mean 12 months, maximum 93 months). Five of the 16 studies presented their findings in subgroups based on age. The mean (standard deviation) size of each of these subgroups was 33.33 (26.69) for the boys and 38.75 (27.10) for the girls. Fifteen studies were conducted in Western countries (the United States, Canada, Europe, or Israel) and one in Hong Kong with children of Chinese ethnicity, and the findings were published between 1980 and 2016. Three studies were conducted in the child's home, ten in a laboratory setting, and three in a nursery. In four studies, the child played alone; in seven, an adult was present but not interacting with the child; and in five studies, there was moderate or full interaction between parent and child. Thirteen studies included gender-neutral toys as well as gender-typed toys in the stimuli presented to children. Fourteen studies were cross-sectional, and two had both cross-sectional and longitudinal elements.

2.5 | Methodological quality

Regarding the quality of the studies, in general, the NOS scores were all of at least moderately good quality. Table 2 shows that the scores ranged from 5 to 8 out of a maximum possible of 9.

3 | RESULTS

3.1 | Data analysis

Heterogeneity was assessed using I^2 and chi-square statistics. A meta-analysis showing tolerable heterogeneity ($I^2 < 30\%$ and chi-square $p > .10$) will be analysed using fixed-effects models, and studies beyond these limits will be analysed using random-effects models.

3.2 | Meta-analyses

The meta-analysis found that boys played with male-typed toys more than girls did (Figure 2a). The effect size of this difference was large (Cohen's $d = 1.03$, $p < .0001$), and girls played with female-typed toys more than boys did (Figure 2b). The effect size of this difference was large (Cohen's $d = -0.91$, $p < .0001$).

In interpreting Figure 2a,b, a minus sign in front of the d value indicates that girls engaged in playing with the toy more than boys did, and a plus sign in front of the d value indicates that boys engaged in playing with the toy more than girls did. Thus, in Figure 2a, showing the results of playing with male-typed toys, where $d = [+].103$, this means that boys played with male-typed toys more than girls did.

There was substantial heterogeneity ($I^2 > 70\%$), so the random-effects model was used. Figure 2a,b and Table 3 show that there were large sex differences in toy preference.

3.3 | Meta-regression

A random-effects meta-regression was performed on the data using Stata's *metareg* command. A random-effects model was chosen not only because of the heterogeneity seen in Table 4 but also because the random-effects model generally reduces the chance of Type 1 error in meta-regression analysis (Harbord & Higgins, 2008). The two criterion variables were sex difference in play with male-typed toys and female-typed toys, and the six predictors were (a) the age of the child in months; (b) the presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); (c) setting of study (1 = home; 2 = laboratory; 3 = nursery); (d) how gender egalitarian the country was at the time the study took place; (e) years since publication; and (f) whether or not gender-neutral as well as gender-typed toys were included in the study.

TABLE 3 Findings of the meta-analysis of 16 studies, with 27 groups of children in total

Toy type	Cohen's d [95% CI]	Z (p)	χ^2 (p)	I^2 (%)
Male-typed toys	1.03 [0.82, 1.24]	9.56 (<.0001)	96.56 (<.0001)	73.1
Female-typed toys	-0.91 [-1.12, -0.40]	-8.32 (<.0001)	104.22 (<.0001)	75.1

Note. Subgroup analyses are based on the type of toy (male- or female-typed) that children played with. The random-effects model was used due to the substantial homogeneity. With the fixed methods model, the d values were smaller for male-typed toys (0.861, 95% CI [0.76, 0.97]) and female-typed toys (-0.759, 95% CI [-0.861, -0.657]). Cohen's d , Z, and chi square have been capped at 0.01.

TABLE 4 Findings of the meta-regression of 16 studies, with 27 groups of children in total

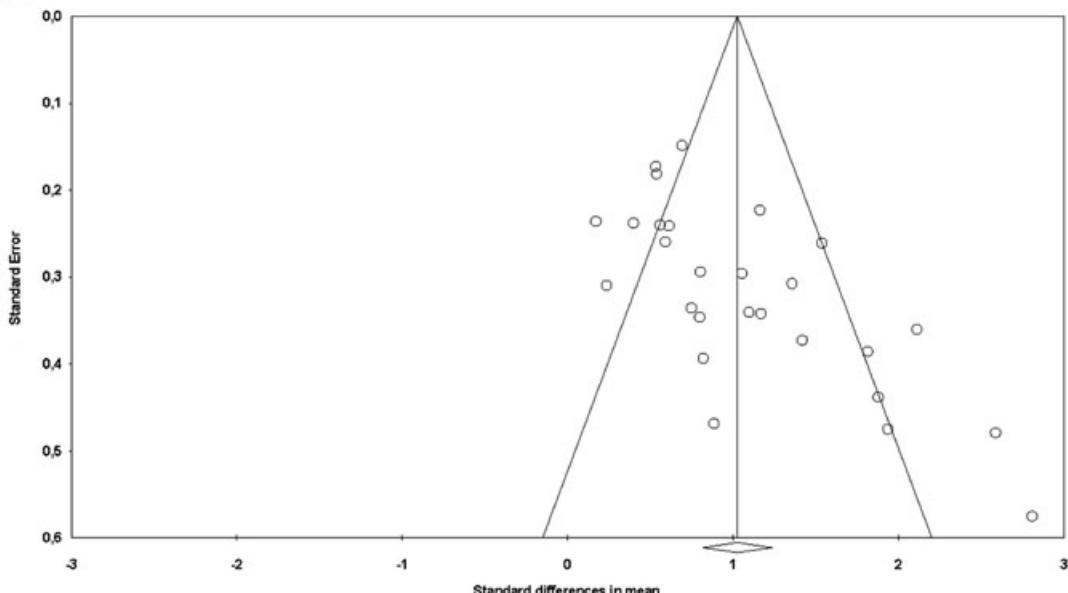
Variable	Male-typed toys			Female-typed toys		
	β	SE β	t	β	SE β	t
Present	.04	0.22	0.19	.09	0.18	0.46
Setting	.14	0.30	0.47	.09	0.28	0.34
Country (Gender Inequality Index)	.07	1.90	0.04	.59	1.70	0.35
Publication years	.01	0.02	0.31	.00	0.02	0.03
Age	.01	0.01	1.77	-.01	0.01	-1.81
Gender neutral toys	-.25	0.47	-0.53	-1.72	0.84	0.477

Note. Subgroup analyses are based on the type of toy (male or female typed) that children played with. Present = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place; Publication years = years since publication of the study; Age = age of group in months; Gender-neutral toys = whether or not gender-neutral as well as gender-typed toys were included in the study as stimuli. β and SE β values have been capped at 0.01.

Table 4 shows that, for male-typed toys, the covariates predicted only 11.51% of the variation in the sex difference in time played with male-typed toys, and the model as a whole did not perform significantly better than chance, $F(6, 19) = 1.30, p = .301$. In this model, none of the predictors were significant. The strongest predictor was age, but was nonsignificant ($t = 1.77, p = .093$), indicating that the sex differences in play with male-typed toys become nonsignificantly larger as children grow older.

For female-typed toys, although the covariates predicted 38.17% of the variation in the sex difference in time played with male-typed toys, the model as a whole did not perform significantly better than chance,

(a)



(b)

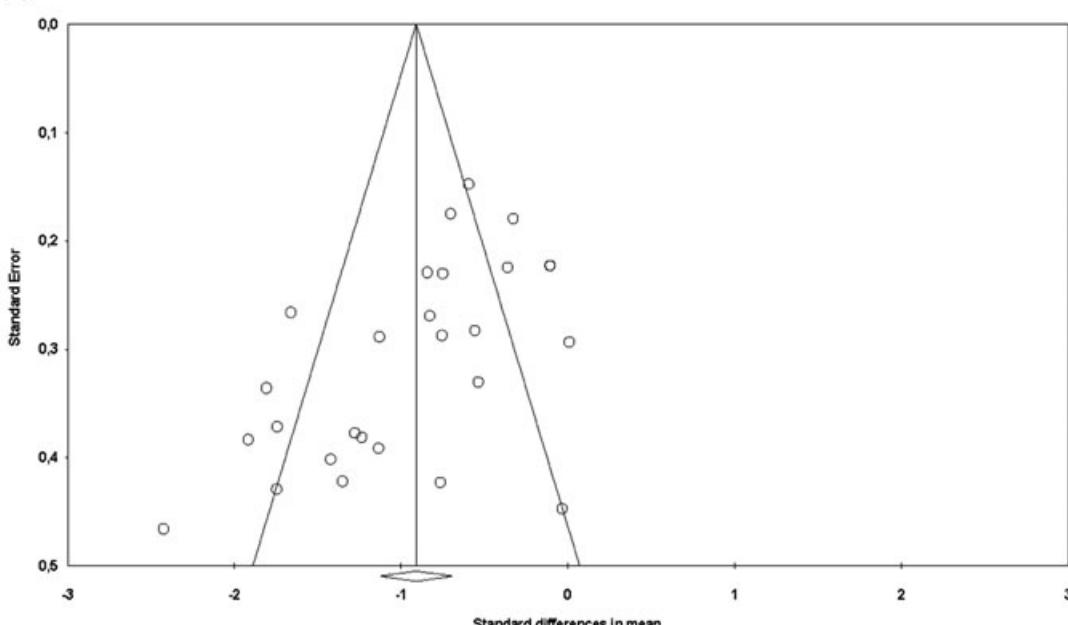


FIGURE 3 Funnel plot of the sex difference in play with (a) male-typed toys and (b) female-typed toys

$F(6, 19) = 2.55, p = .056$. In this model, none of the predictors were significant. The strongest predictor was age but was nonsignificant ($t = -1.81, p = .086$), indicating that the sex difference in play with female-typed toys become non-significantly smaller as children grow older.

3.4 | Publication bias

To assess publication bias, funnel plots were created (Figure 3a,b). Overall, most of the studies are within the 95% confidence intervals, despite the fact that study heterogeneity will tend to place the studies beyond these confidence intervals. However, asymmetry is in evidence. The funnel plot of the findings for the sex differences in play with male-typed toys (Figure 3a) is asymmetrical (seen in the tail in the lower right of the plot). The funnel plots for studies with female-typed toys (Figure 3b) is similar to Figure 3a, except that the trend in the effect size was in the opposite direction (i.e., the tail is in the lower left) as would be expected from the forest plots in Figure 2a,b. The three main outliers in Figure 3a (from right to left) are Pasterski et al. (2005; with father present), Todd et al. (2016; oldest group), and Pasterski et al. (child alone), and the four main outliers in Figure 3b are Todd et al. (oldest group; youngest group), Pasterski et al. (child alone), Arthur (2014), and Doering et al. (1989). Often, asymmetry is caused by publication bias due to small study effects, that is, a bias towards small studies finding large effects and being published. Although many of the groups in this meta-analysis are, technically, small ($N < 50$), the outlying groups are not small relative to nonoutlying groups. However, what the outlying groups all have in common is that they represent either older age groups or younger groups than those represented by nonoutlying groups. The NOS scores are not implicated because Table 2 shows that the outlying groups all had scores around the midrange, either 6 or 7 from a NOS score range of 5 to 8 (Doering et al., 1989 = 6; Pasterski et al., 2005 = 7; Todd et al., 2016 = 7).

3.5 | Analysis of percentage of time played with male-typed toys by boys and girls separately

The meta-regressions did not find any significant predictors of sex difference in toy play. However, in bivariate correlations between toy play and predictor variables, some of the predictors showed different patterns of correlations when data from boys and girls were analysed separately. Such differences violate the assumptions for linear regression (homogeneity of regression slopes), and combining data from boys and girls, as in the above meta-regressions, is prone to Type 2 error and may result in null findings. For this reason, in the following section, the regressions are performed separately for boys and girls. The outcome in these regressions cannot be sex differences in toy play

TABLE 5 Spearman's (bivariate) correlations between eligible predictors and the percentage of time played with male-typed toys for boys ($N = 27$) and girls ($N = 27$)

Gender	Present	Country (GII)	Publication years	Age	Home versus laboratory ^a	Nursery versus laboratory ^a	Gender-neutral toys
Boys	-.17	-.03	.47*	.68****	-.46*	.18	-.13
Girls	.14	-.02	.70****	.17	-.23	-.07	.17

Note. Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place; Publication years = years since publication of the study; Age = age of group in months; Gender-neutral toys = whether or not gender-neutral as well as gender-typed toys were included in the study as stimuli. R_s has been capped at 0.01.

^aIn order to not violate the assumptions of the regression, the categorical data were split.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

**** $p < .0001$ (two-tailed).

(because toy play cannot be compared by sex in an analysis where boys and girls are assessed separately), so the outcome variable in the following regressions is the percentage of time played with male-typed toys relative to female-typed toys. Thus, although in the previous section the toy preference was analysed separately for each toy type (combining data from boys and girls), in the following section, toy preference is analysed separately for boys and girls (combining male- and female-typed toys).

Table 5 shows the bivariate correlations between predictors and the percentage of time played with male-typed toys for boys and girls.

Although Table 5 shows positive correlations for both boys and girls, further analysis reveals other differences. Figure 4a,b shows, respectively, scatterplots of the relation between age and percentage of time played with male-typed toys for boys (Figure 4a) and age and percentage of time played with male-typed toys for girls (Figure 4b). Figure 4a shows a positive correlation, indicating that boys play more with male-typed toys as they grow older ($r_s = .43$, $n = 27$, $p < .0001$, two-tailed). In contrast, Figure 4b shows a negative quadratic correlation, indicating that girls' play

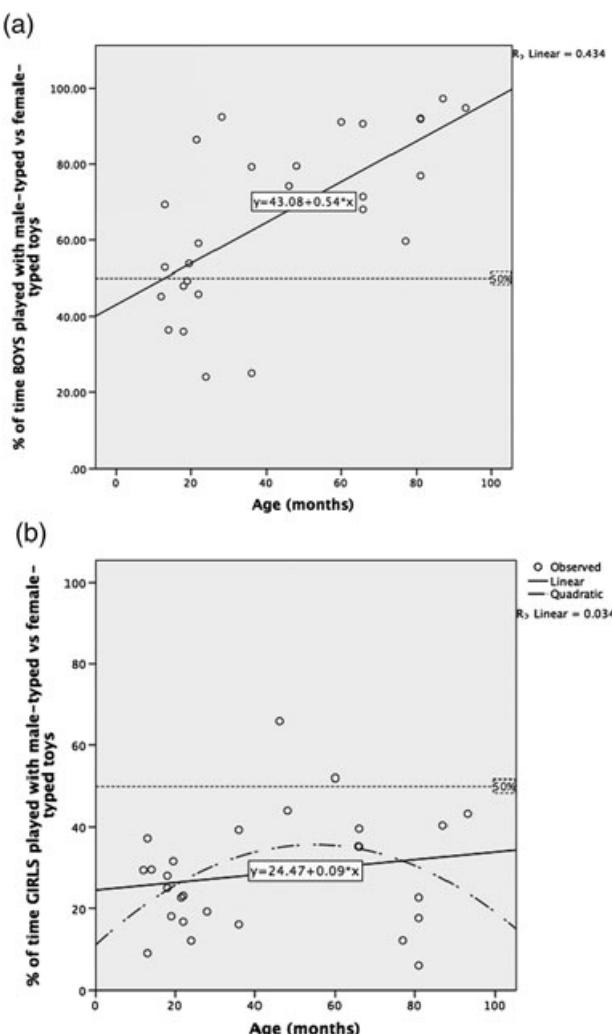


FIGURE 4 Scatterplot of the percentage of (a) time boys played with male-typed toys relative to female-typed toys, by age, and (b) time girls played with male-typed toys, by age

with male-typed toys increases to a peak at about 4 years old and then reduces thereafter. This quadratic trend was not significant using linear correlations ($r_s = .03, n = 27, p = .357$, two-tailed) but was closer to significance using the SPSS curve fit analysis for quadratic functions ($R^2 = .11, r = .34, n = 27, p = .223$, two-tailed).

Figure 5a and 5b shows, respectively, scatterplots of the relation between including gender-neutral toys in the stimuli and the percentage of time played with male-typed toys for boys (Figure 5a) and girls (Figure 5b). Figure 5a shows that boys play nonsignificantly less with male-typed toys when gender-neutral toys are included ($r_s = -.13, n = 27, p = .422$, two-tailed), and Figure 5b shows that girls play nonsignificantly more with male-typed toys when gender-neutral toys are included ($r_s = .17, n = 27, p = .272$, two-tailed).

Figure 6a,b shows, respectively, scatterplots of the relation between testing children at home and testing children in a laboratory for boys (Figure 6a) and girls (Figure 6b). Table 5 shows that boys play significantly less with male-typed

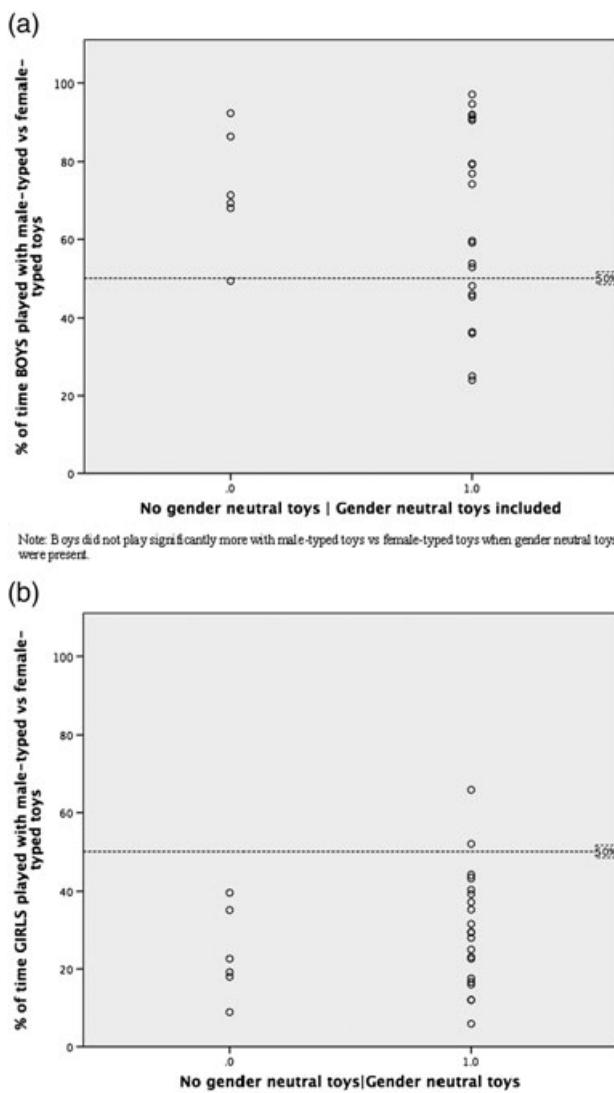


FIGURE 5 Scatterplot of relation between including gender-neutral toys in the stimuli and percentage of time played with male-typed toys (a) for boys and (b) for girls

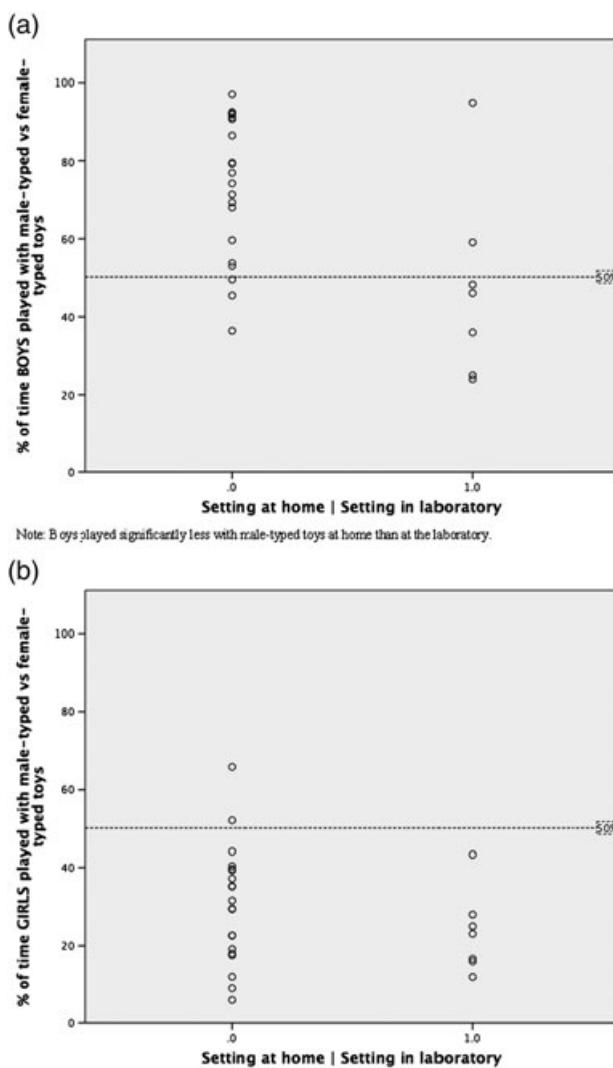


FIGURE 6 Scatterplots of the relation between testing children at home and testing children in a laboratory (a) for boys and (b) for girls

toys at home ($r_s = -.46, n = 27, p = .017$, two-tailed), whereas girls play with male-typed toys nonsignificantly more in a home setting than in a laboratory ($r_s = -.23, n = 27, p = .253$, two-tailed).

Because of these differences, these variables were analysed separately for boys and girls in multiple linear regression, using SPSS 22.

Table 6 shows the results of the multiple linear regressions of the predictors, with percentage of time played with male-typed toys relative to female-typed toys as the dependent variable. Age was a significant predictor for boys, indicating that older boys played more with male-typed toys relative to female-typed toys than younger boys did. The length of time since publication of the study was a significant predictor for girls, showing that girls play less with male-typed toys and female-typed toys in more recent studies. The regression model for boys was a strong predictor of play, adjusted $R^2 = .57, F(7, 18) = 5.69, p < .001$. For girls only, the regression model was a moderate predictor of play, adjusted $R^2 = .43, F(7, 18) = 3.68, p = .012$.

TABLE 6 Multiple linear regression for boys (left side) and girls (right side), with percentage of time played with male-typed toys relative to female-typed toys as the dependent variable

Variable	Boys			Girls		
	B	SE β	β	B	SE β	β
Present	2.2	5.3	.06	2.1	3.8	.10
Setting ^a						
Home versus laboratory	-13.73	9.85	-.27	-.51	5.81	.50
Nursery versus laboratory	6.77	11.66	.13	16.26	8.04	.50
Country (gender inequality index)	6.30	42.94	.02	1.16	1.64	.49
Publication year	0.59	0.48	.24	1.14*	0.33	.77
Age	0.46	0.14	.55	-0.06	0.10	-.12
Gender neutral toys	-9.20	14.45	-.17	7.16	10.32	.22

Note. Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place; Publication years = years since publication of the study; Age = age of group in months; Gender-neutral toys = whether or not gender-neutral as well as gender-typed toys were included in the study as stimuli. β and SE β values have been capped at 0.01.

^aIn order to not violate the assumptions of the regression, the categorical data were split.

* $p < .05$,

** $p < .01$ (two-tailed).

3.6 | Analysis of raw seconds of play with each toy type by boys and girls separately

In the Section 3.2, toy preference was analysed separately for each toy type (combining boys' and girls' data), and in Section 3.5, toy preference was analysed separately for boys and girls (combining male- and female-typed toys). In this section, the actual duration of play with each toy type was assessed for boys and girls separately. Thus, the number of seconds of play with male- and female-typed toys, rather than the percentage of play with male-typed toys relative to female-typed toys, was assessed. Table 7 shows the Spearman bivariate correlations for both boys' and girls' play behaviour for male- and female-typed toys. It was found that for boys, age correlates significantly with the raw seconds of play with male-typed toys ($r = .54$, $n = 27$, $p < .05$), whereas play with female-typed toys negatively correlates

TABLE 7 Spearman bivariate correlations for predictors and boys and girls play behaviour with male-typed toys and female-typed toys

	Present	Country (gender inequality index)	Publication years	Age	Home versus laboratory	Nursery versus laboratory	Gender-neutral toys
Boys	Male typed	.09	.17	.71**	.54*	-.34	.38
	Female typed	.41*	.09	.14	-.51*	.04	.25
Girls	Male typed	.17	-.03	.65**	.26	-.27	.32
	Female typed	.06	-.02	.50*	.20	-.24	.32

Note. Presence = presence of parent (1 = absent; 2 = minimal interaction; 3 = moderate or full interaction); Setting = location of study (1 = home; 2 = laboratory; 3 = nursery); Country = gender inequality index, a measure of how gender egalitarian the country was at the time the study took place; Publication years = years since publication of the study; Age = age of group in months; Gender-neutral toys = whether or not gender-neutral as well as gender-typed toys were included in the study as stimuli.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

**** $p < .0001$ (two-tailed).

($r = -.51, n = 27, p < .05$). Additionally, it was found that for girls, the older the study, the longer they played with both male-typed ($r = .65, n = 27, p < .001$) and female-typed toys ($r = .50, n = 27, p < .005$). For boys, the older the study, the longer they play with male-typed toys ($r = .71, n = 27, p < .0005$), but there was no effect of time on their play with female-typed toys ($r = .05, n = 27, p = .816$).

4 | DISCUSSION

This meta-analysis found sex differences in the preferences for male- and female-typed toys among boys and girls aged between 1 and 8 years in the predicted directions; boys played with male-typed toys more than girls did, and girls played with female-typed toys more than boys did. The effect sizes of these differences were large (Cohen's $d = 1.03$ and $d = -0.91$, respectively). This finding, from 16 observational studies measuring boys' and girls' independent choice of gender-typed toys in free play situations, is consistent with the wider literature on sex differences in play; both observational and questionnaire data indicate that typically developing boys and girls prefer different toys (Hines, 2004). Meta-regression detected no significant effect of the presence or absence of an adult, the study setting, the gender equality status of the country, year of publication, and presence of gender-neutral toys, indicating a consistent effect of child sex on toy preference across a range of environmental circumstances. However, further analysis revealed significant effects of age and indicated that boys' and girls' preferences develop differently with age and may be subject to differential change over contexts and across historical time.

4.1 | The effect of age

The included studies of children's independent behaviour varied by social, geographical, and historical contexts and by stimuli representative of toy type. Although the meta-regressions reported in Section 3.3 found no effect of age, this may have been due to nonlinearity in the data for girls (Figure 4b). In Section 3.5, the analysis of percentage of time played with male-typed toys relative to female-typed toys by boys and girls separately indicates that boys and girls show different developmental patterns with respect to toy preference. For boys, there is a clear positive correlation between time playing with male-typed toys and age, but the findings for girls do not show a mirror image of that profile. For girls, there is a clear negative quadratic (inverted U-shaped) correlation with age. Figure 4a suggests that younger boys (up to ~20 months) play more with female-typed toys, whereas Figure 4b suggests that girls' interest in male-typed toys peaks around four years and then tails off somewhat. Although Table 5 shows that this finding for girls was nonsignificant, the significance of the trend cannot be detected by linear regression because the trend is nonlinear. Our findings, therefore, are consistent with an effect of cognitive development and differential socialization on the strength of boys' and girls' preferences for gender-typed toys.

These findings may indicate the differential effects of social experiences on boys' and girls' cognitive development. Infants' application of rigid stereotypes may be consequential on the acquisition of a gendered identity and their accumulation of knowledge about typical gendered behaviour in others (Zosuls et al., 2009). This view is consistent with our finding of an increase in the magnitude of boys' and girls' free preference for toys typed to their own gender across this age period. After infancy, social influences (from parents, peers, educators, and the media) are more accessible to children and are likely to differ, by gender, in both focus and intensity, according to prevailing norms of gendered attitudes and behaviour. Children's stereotypical beliefs increase from the ages of 3 to 5 years (Halim, Ruble, Tamis-LeMonda, & Shrout, 2013), peak between the ages of 5–7 years, and become more flexible thereafter (Trautner et al., 2005), perhaps because children gain an appreciation of individual differences in preferences. The trend for age-related changes identified in this meta-analysis fits the view that the strength of stereotypes will "wax and wane across development" (Martin & Ruble, 2004, p. 68) but indicates that transition away from greater rigidity may begin earlier than suggested.

A less flexible application of stereotypes in younger children is indicated by the finding that 3- and 5-year-old children ridicule same-sex peers who play with "cross-sex" toys (Langlois & Downs, 1980) and this tendency possibly persists longer for boys. The oldest children in the included studies were aged about 8 years, which may have been too early to detect a reduction in boys' stereotyped behaviour. Our findings indicate that reduction in gender-typed toy preference may apply only, or more, to girls. This change is consistent with the finding from a questionnaire study that, as age increased, both girls and boys increasingly requested male-typed toys as gifts (Etaugh & Liss, 1992).

Of the four included studies reporting age comparisons, one found that both boys and girls played longer with female-typed toys as age increased (Schau et al., 1980) and another that boys' play time with a truck and doll were relatively consistent across the ages of 24 and 36 months, whereas girls' play time with both of these toys increased slightly with age (Zosuls et al., 2014). The others show that female-typed toys became less interesting to girls as well as boys with age (Servin et al., 1999; Todd et al., 2016). One possible explanation for the inconsistency might be differences in the selected stimuli; for example, Schau et al. (1980) included both a doll house and a mixer as female-typed toys, both of which were popular with boys.

4.2 | Presence or absence of an adult and setting of study

Contrasting studies or conditions where children played alone or in interaction with a parent have the potential to compare the immediate effect of social influence and individual play, which might better reflect the child's intrinsic preferences. In the meta-regression, the immediate social context of testing (child playing alone or in the presence of or in interaction with a parent) did not predict the size of sex differences in play with male- or female-typed toys. Of the included studies that compare infants across conditions in which the same participant plays (a) alone and (b) in interaction with a parent, findings are inconsistent. One study found stable sex differences across play-alone and play-with-parent conditions (Alexander & Saenz, 2012), whereas others found that infants played with gender-typed toys less when with their mother than when alone (Zosuls et al., 2009) and more with male-typed toys when with their father than when alone (Servin et al., 1999). In studies involving older children, no differences were found between play-alone or play-with-parent conditions for 3–5-year-olds (Servin et al., 1999) but typically developing 3–10-year-old girls were found to play more with female-typed toys when alone than when with either parent (Pasterski et al., 2005). These findings have implications for the design of studies of sex differences in toy preference and may indicate nuances in the immediate social influences on boys and girls of different ages. Typically, fathers are underrepresented in studies that include play with a parent, and this may influence the results when sex of parent is not controlled for.

In the majority of studies, an adult observer (often the researcher) is present. Conditions described as play alone do not necessarily preclude the presence of an inactive parent or adult researcher (especially with younger participants). The effect of such variables could be explored further. Schau et al. (1980) found no effects of adult observer presence or absence or observer sex on typically developing preschool children's gender-typed toy selection. However, an effect of adult presence may vary by age or developmental stage; preschool-aged boys and girls rated as gender aschematic displayed more gender-typical play behaviour in the presence of an adult observer than when they were alone, whereas no such difference was found for children rated as gender schematic (Wilansky-Traynor & Lobel, 2008). Taken together, findings relating to social context indicate a potential interplay of factors, including the sex of the play partner or observer and the child's sex and developmental stage and the level and focus of interaction between adult and child. In modern Western societies, at least, there may be a move towards parental encouragement of gender-neutral play (see Zosuls et al., 2009). Studies that strictly control for immediate social influences on children at different ages may have the potential to inform our understanding of how expectations of social behaviour are modified by specific contexts, including self-presentation in the presence of other children.

Study locations (home, laboratory, nursery) may also impact children's sex-typed behaviour. Single studies of sex differences in children's free play have not contrasted study settings, but the meta-analysis provides an opportunity to do so. Interestingly, as Table 5 shows, boys played with male-typed toys less when at home than when in a laboratory

setting ($r_s = -.46, p < .05$); however, although the trend is in the same direction, the finding for girls is nonsignificant ($r_s = -.23, p > .05$). It is difficult to make conclusions from this finding, but the relative formality and a potential testing element of the context may draw children, especially boys, to toys that they are aware are typed to their own gender.

4.3 | The effect of country (gender egalitarianism) in which the study was conducted

The opportunity to make comparisons of children's toy preferences across different countries allows a test of wider social influence on both children and their families. Although the influence of adults and of the media is likely to vary by cultural context, the size of sex differences in children's preferences for male- and female-typed toys did not appear to be smaller in studies conducted in more egalitarian countries; large effects of sex were apparent across regions where the GII varied. It could be argued that nonsignificant findings were due to low power. Although the number of groups was small, the number of participants was about 26 times greater than the number of groups. Thus, although technically the regressions were of low statistical power, the β values may generalize well to the general population of children, and relatively large beta values (e.g., .59 for GII for girls) should be deemed as worth consideration, even if not statistically significant.

There is probably no perfect way to measure gender egalitarianism in any country at any given time, and subsections of the country-wide population are likely to vary on the included measures. The GII has an advantage over some other indexes in having scores for many countries going back to 1995. This is not ideal for the range of study dates in this meta-analysis but is better than the gender gap index, another popular tool, which goes back only to 2006. The GII is only produced every few years, so the year of study publication and year GII was conducted cannot be matched exactly. However, only the three earliest studies were published more than 2 years prior to the dates of GII reports, and these do not vary much within countries over time (e.g., the GII in the United Kingdom in each of the five years measured: 2000 = 0.23; 2005 = 0.22; 2010 = 0.21; 2012 = 0.21; 2013 = 0.19). Despite the relative consistency across regions showing some support for biological determinants of preferences, a limitation of this meta-analysis, and of this research area in general, is that most of the research has been conducted in Western countries and findings cannot be generalized elsewhere.

4.4 | The effect of the year the study was published

Table 5 shows that, in bivariate correlation, the year of publication of the study was a significant predictor for boys ($r_s = .47, p < .05$), and more so for girls ($r_s = .70, p < .0001$). Table 5 shows that this effect remains significant when the effect of other variables is accounted for. Girls play significantly less with female-typed toys and male-typed toys in more recent studies. Therefore, the findings of this study indicate a possible effect of historical time on toy preference. This finding may indicate moves towards greater gender equality in Western societies (Schwartz & Rubel-Lifschitz, 2009). One interpretation is that, over the years, girls have been exposed to increasing social pressure to play with neutral-typed toys as the volume of advertising to children and expansion in the number and type of children's media outlets in the United States, and doubtless in other developed countries, have increased enormously over time (Bakir, Blodgett, & Rose, 2008). A further possible explanation would be that female-typed toys have become less interesting or attractive to girls over time, yet the included studies have typically and fairly consistently used dolls, cosmetics, and kitchen equipment as female-typed toys.

A possibility to be considered relates to differences in the researchers' selections of male-typed toy stimuli at in different time periods. Earlier studies included in the meta-analysis are more likely to offer toys associated with aggression, for example, a dart gun and army shirt (Doering et al., 1989) and fighting figures (Servin et al., 1999), yet a gun was used as a male-typed toy in one recent study (Li & Wong, 2016). Of course, other toys, not necessarily male-typed, can afford aggressive play, as Schau et al. (1980) note in their observation that boys played with a toy food mixer as if it were a machine gun. It is difficult to be precise about the year in which a study was conducted

as there is typically some time lag between this and the date of the resulting publication. The finding that girls spend a relatively shorter time playing with gender-typed toys in recent studies is of interest, given the indication that some parents and educators may currently promote gender-neutral toys for girls. The number and specific characteristics of toys by study are not always reported, yet no systematic changes over time are apparent from the information available. However, media coverage, cultural trends, or current events may affect children's toy preferences. For example, at the time when Benjamin (1932) found that girls played with a toy aeroplane for similar times to boys, female aviators were in the news; Amy Johnson flew from England to Australia in 1930, and Amelia Earhart flew solo across the Atlantic in 1932.

4.5 | The effect of inclusion of gender-neutral toys

Like age, the inclusion of gender-neutral as well as gender-typed toys as stimuli appeared to be associated with opposite (though statistically nonsignificant) relationships for boys and girls; boys played less with male-typed toys when gender-neutral toys were included in the stimulus materials (Figure 5a), and girls played more with male-typed toys when gender-neutral toys are included (Figure 5b). However, in both cases, although the directions of correlations remained the same in multiple regression, the apparent effect remained nonsignificant when the other variables were taken into account (Table 5). However, the opposite pattern in boys and girls is interesting and may have implications for the design of future studies of sex differences in children's toy preferences, with the implication that some toys classified as gender neutral may appeal more to boys than to girls.

In general, the studies included in the meta-analysis varied with respect to the number of toys chosen as representative of each toy-type category, the degree to which they are stereotyped to one gender and the duration of play sessions. Only three studies did not include gender-neutral toys, and so no comparison of effect sizes between these and the other studies was conducted.

4.6 | Limitations

One problem facing researchers is the choice of toys to include as gender typed or gender neutral and the number of toys offered. The most popular gender-typed toys are dolls and vehicles, yet each of these comes in many forms; for example, a doll can represent a baby, a male or female child, or a male or female adult and be soft or rigid in composition and vary in colour. The suitability and attractiveness of toys are likely to differ with child age, and some studies use the same stimuli across a relatively wide age range. In one study, the amount of time parents spent playing with specific toys in interactions with their children (aged 27–64 months) indicated that female-typed toys were the least entertaining (Idle et al., 1993), and this factor may have the potential to bias results.

In many individual studies, the selection of stimuli is not discussed. However, one reported solution has been to use toys that have elicited gender differences in previous studies (e.g., Lamminmäki et al., 2012; Zosuls et al., 2014) or have been classified as masculine or feminine by parents and nonparents (e.g., Todd et al., 2016; van de Beek et al., 2009). Toys typically selected as neutral include books and puzzles, but there is some overlap between categories; for example, Turner, Gervai, and Hinde (1993) define a jigsaw puzzle as female typed. Perhaps the most thorough consideration of the choice of stimuli comes from Idle et al. (1993); in this study, parents were asked to sort the stimuli toys into masculine, feminine, and neutral categories. Interestingly, both mothers and fathers categorized many more toys as neutral than the researchers did. However, despite methodological variation between observational studies in the choice and number of stimuli offered, context of testing, and age of child, the general consistency in finding sex differences in children's preferences for toys typed to their own gender indicates the strength of this phenomenon.

There was a substantial heterogeneity in findings from the included studies—roughly 70% for both male- and female-typed toys. Although any effects of this were mitigated by using the random-effects model, we note that the fixed-methods model produced slightly smaller Cohen's d values for male-typed toys (1.03, 95% CI [0.82, 1.24]) and female-

typed toys (-0.91 , 95% CI $[-1.12, -0.40]$). On the whole, it would probably be reasonable to suggest that the effect size is large for the sex difference in preference for male-typed toys and is moderate to large for female-typed toys.

Some studies of children's free preference for toys (e.g., O'Brien & Huston, 1985; Roopnarine, 1986) were excluded from the meta-analysis because reports did not include essential information; for example, standard deviations were not reported and could not be obtained from the authors; these tended to be earlier studies, which may have affected interpretations based on the effect of year of publication.

This meta-analysis comprised papers relating to children's free preferences for gender-typed toys, and some did not meet the criteria for inclusion specified in the Section 2. Many of the excluded studies showed interesting results, largely consistent with the findings of the included papers. For example, cross-cultural studies are particularly informative on the effect of culture; a study conducted in Hungary and the United Kingdom (Turner et al., 1993) was excluded because children were in interaction with other children. The impact of hormonal variation on the toy preferences of typically developing children allows exploration of the biological explanation; a German study measured the association between digit ratio (2D:4D), as a measure of early testosterone exposure, and sex-typed play but used a measure completed by parents rather than the observational method (Hönekopp & Thierfelder, 2009).

It could be argued that a bias to publish only positive findings of sex difference in toy choice is unlikely as a null finding in any age group or context would be of interest because of the social and political implications of stereotyping. Some nontypical preferences are reported in studies that did not meet the inclusion criteria of this meta-analysis; for example, at 12 months, boys played with a vacuum cleaner more than girls and no significant difference was found in the time boys and girls played with a truck or a shovel (Snow, Jacklin, & Maccoby, 1983). In the included studies, as stated above, some atypical results were found, and, of course, there was considerable variation in individual children's preferences.

There was no support for the hypothesis that the presence of adults will affect the magnitude of preference for gender-typed toys. However, fathers, and possibly male observers, are underrepresented in studies comparing boys' and girls' toy preferences in specific social contexts, and this may limit the interpretation with respect to this factor.

One important point to note is that all but one of the studies included in this meta-analysis and, indeed, most studies of children's toy preferences are conducted in Western countries. It is important therefore to extend toy preference research to other geographical locations.

4.7 | Methodological concerns

It should be noted that the GII is a blunt measure of societal attitudes in a particular country; it is likely that environmental influences may vary more by regional and cultural subgroup norms within countries than between some of the countries where the included studies were conducted.

In general, the typicality of the participant samples can be questioned; participants' families may typically represent comparatively high socio-economic groups and therefore be associated with nontraditional sex-typed attitudes (Zosuls et al., 2009). Nevertheless, this is not reflected in the findings of large sex differences in toy preferences in the meta-analysis. Studies involving relatives of children recruited in clinical contexts are likely to be drawn from the broader population (Berenbaum & Hines, 1992; Berenbaum & Snyder, 1995; Doering et al., 1989; Pasterski et al., 2005), as are studies conducted in multicultural nurseries (Todd et al., 2016). One possible caution is that parents may encourage gender-stereotyped play more outside a research setting.

4.8 | Directions for future research

Key recommendations for future research include the following:

1. We recommend attention to age of child when making comparisons between typically developing children and those with atypical levels of hormones or other specific characteristics.

2. We argue for the value of longer term longitudinal studies in addressing the relationship between boys' and girls' toy preferences and the development of gendered identity. However, such studies are compromised by repeated exposure to same or similar toys, suitability of toys for the different age groups, and the relative attractiveness of toys (Lobel & Menashri, 1993).
3. We suggest that randomized controlled trials might be useful in testing for: geographical variation; typical and atypical children; and differences in behaviour across contexts of testing, for example, study locations and social contexts. Where younger children are tested playing alone, ethical and practical constraints prevent researchers from isolating a young child completely, and it may be more useful to test between play alone and play in social contexts in a safe and familiar home environment.
4. We suggest attention to selection of test stimuli, particularly to the inclusion or noninclusion and characteristics of toys classified as gender neutral. Meta-regression reveals that the inclusion of gender-neutral toys may have a differential impact (albeit statistically nonsignificantly) on boys' and girls' toy preferences, with boys playing less with male-typed toys when gender-neutral toys are included and girls playing more with male-typed toys when gender-neutral toys are included. Therefore, future studies could include a condition where gender-neutral toys are present and a condition where they are absent.

Methodology might be improved by, for example, greater attention to the choice of stimuli, as exemplified in Idle et al. (1993), a study with a high NOS methodological quality score. Such attention might shed light on the changing perceptions and behaviours of parents and children across a range of ages. Also, an extension of studies to nonhuman primates might contribute to evidence related to biological theories.

5 | CONCLUSION

In view of the societal interest in and clinical implications of sex-typed behaviour, it is important to determine its origins: the present review analysed studies of boys' and girls' (aged between 1 and 8 years) free preference for gender-typed toys, asking questions related to the degree to which sex, age, and various social factors impacted on their behaviour. In general, the finding of robust sex differences in boys' and girls' toy preferences across a range of ages, different time periods, countries, and settings indicates an innate influence on this behaviour, an effect that appears to be subject to modification by developmental and social factors arising at different ages.

In observational studies of the type reviewed here, it is impossible to determine the degree to which findings are a result of biological predisposition or environment. However, this study attempted to assess the variables most likely to be indicative of either nature or nurture and must conclude that there is some (circumstantial) evidence for both sides of the argument. When assessing the effect of publication year, we suggest that there is more evidence of the effect of environment on girls' than on boys' toy preferences. In the assessment of the effects of age, it could be argued that social effects on boys are stronger or persist longer than those on girls or, alternatively, that there is a stronger biological predisposition for boys' attraction to particular object features. However, these suggestions are necessarily speculative, and we recommend that experimental evidence is required in order to shed light on the specific questions that are identified in the meta-analysis, including the nuances of social influence.

Vygotsky (1978) regarded toys as cultural artefacts or "tools of the mind" that are designed with specific skills, either cognitive or social, in mind. One implication of the findings of the meta-analysis is that it may be important to ensure that toys that are attractive to each sex are not restrictive in the skills that they afford.

ORCID

Brenda K. Todd  <http://orcid.org/0000-0003-0098-5623>

Rico A. Fischer  <http://orcid.org/0000-0002-6966-0803>

John A. Barry  <http://orcid.org/0000-0002-8931-0333>

REFERENCES

- Alexander, G. M. (2014). Postnatal testosterone concentrations and male social development. *Frontiers in Endocrinology*, 5, 15. <https://doi.org/10.3389/fendo.2014.00015>
- Alexander, G. M., & Hines, M. (2002). Sex differences in response to children's toys in nonhuman primates (*Cercopithecus aethiops sabaeus*). *Evolution and Human Behavior*, 23, 467–479.
- Alexander, G. M., & Saenz, J. (2012). Early androgens, activity levels and toy choices of children in the second year of life. *Hormones and Behavior*, 62(4), 500–504.
- Alexander, G. M., Wilcox, T., & Woods, R. (2009). Sex differences in infants' visual interest in toys. *Archives of Sexual Behavior*, 38(3), 427–433. <https://doi.org/10.1007/s10508-008-9430-1>
- Arthur, A. E. (2014). *Does social categorization affect toddlers' play preferences? An experimental test* (MA dissertation). Available from Dissertation Abstracts International: Section B, The Sciences and Engineering, 74(11-B(E)).
- Bakir, A., Blodgett, J. G., & Rose, G. M. (2008). Children's responses to gender-role stereotyped advertisements. *Journal of Advertising Research*, 48(2), 255–266.
- Banerjee, R., & Lintern, V. (2000). Boys will be boys: The effect of social evaluation concerns on gender-typing. *Social Development*, 9(3), 397–408.
- Benenson, J. F., Tennyson, R., & Wrangham, R. W. (2011). Male more than female infants imitate propulsive motion. *Cognition*, 121, 262–267.
- Benjamin, H. (1932). Age and sex differences in the toy-preferences of young children. *The Pedagogical Seminary and Journal of Genetic Psychology*, 41(2), 417–429.
- Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science*, 3(3), 203–206.
- Berenbaum, S. A., & Snyder, E. (1995). Early hormonal influences on childhood sex-typed activity and playmate preferences: Implications for the development of sexual orientation. *Developmental Psychology*, 31(1), 31.
- Blakemore, J. E., LaRue, A. A., & Olejnik, A. B. (1979). Sex-appropriate toy preference and the ability to conceptualize toys as sex-role related. *Developmental Psychology*, 15(3), 339.
- Campbell, D. W., & Eaton, W. O. (1999). Sex differences in the activity level of infants. *Infant and Child Development*, 8(1), 1–17.
- Cohen-Bendahan, C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: Methods and findings. *Neuroscience & Biobehavioral Reviews*, 29(2), 353–384.
- Collaer, M. L., & Hines, M. (1995). Human behavioral sex differences: A role for gonadal hormones during early development? *Psychological Bulletin*, 118(1), 55.
- Diekman, A. B., & Murnen, S. K. (2004). Learning to be little women and little men: The inequitable gender equality of non-sexist children's literature. *Sex Roles*, 50(5–6), 373–385.
- Doering, R. W., Zucker, K. J., Bradley, S. J., & MacIntyre, R. B. (1989). Effects of neutral toys on sex-typed play in children with gender identity disorder. *Journal of Abnormal Child Psychology*, 17(5), 563–574.
- Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: A meta-analysis. *Psychological Bulletin*, 132(1), 33.
- Etaugh, C., & Liss, M. B. (1992). Home, school, and playroom: Training grounds for adult gender roles. *Sex Roles*, 26(3–4), 129–147.
- Fabes, R. A., Martin, C. L., & Hanish, L. D. (2003). Young children's play qualities in same-, other-, and mixed-sex peer groups. *Child Development*, 74(3), 921–932.
- Fagot, B. I. (1974). Sex differences in toddlers' behavior and parental reaction. *Developmental Psychology*, 10(4), 554.
- Fagot, B. I., & Leinbach, M. D. (1989). The young child's gender schema: Environmental input, internal organization. *Child Development*, 60(3), 663–672.
- Ferrant, G. (2010). The gender inequalities index (GII) as a new way to measure gender inequalities in developing countries. Retrieved from <https://halshs.archives-ouvertes.fr/halshs-00462463>
- Halim, M. L., Ruble, D., Tamis-LeMonda, C., & Shrout, P. E. (2013). Rigidity in gender-typed behaviors in early childhood: A longitudinal study of ethnic minority children. *Child Development*, 84, 1269–1284. <https://doi.org/10.1111/cdev.12057>
- Harbord, R. M., & Higgins, J. P. (2008). Meta-regression in Stata. *Meta*, 8(4), 493–519.
- Hassett, J. M., Siebert, E. R., & Wallen, K. (2008). Sex differences in rhesus monkey toy preferences parallel those of children. *Hormones and Behavior*, 54, 359–364.
- Hines, M. (2004). *Brain gender*. Oxford, UK: Oxford University Press.

- Hines, M., Brook, C., & Conway, G. S. (2004). Androgen and psychosexual development: Core gender identity, sexual orientation, and recalled childhood gender role behavior in women and men with congenital adrenal hyperplasia (CAH). *Journal of Sex Research*, 41(1), 75–81.
- Hines, M., & Kaufman, F. R. (1994). Androgen and the development of human sex-typical behavior: Rough-and-tumble play and sex of preferred playmates in children with congenital adrenal hyperplasia (CAH). *Child Development*, 65(4), 1042–1053. <https://doi.org/10.1111/j.1467-8624.1994.tb00801.x>
- Hönekopp, J., & Thierfelder, C. (2009). Relationships between digit ratio (2D:4D) and sex-typed play behavior in pre-school children. *Personality and Individual Differences*, 47(7), 706–710. <https://doi.org/10.1016/j.paid.2009.06.007>
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581.
- Idle, T., Wood, E., & Desmarais, S. (1993). Gender role socialization in toy play situations: Mothers and fathers with their sons and daughters. *Sex Roles*, 28(11–12), 679.
- Karraker, K. H., Vogel, D. A., & Lake, M. A. (1995). Parents' gender-stereotyped perceptions of newborns: The eye of the beholder revisited. *Sex Roles*, 33(9–10), 687–701.
- Lamminmäki, A., Hines, M., Kuiri-Hänninen, T., Kilpeläinen, L., Dunkel, L., & Sankilampi, U. (2012). Testosterone measured in infancy predicts subsequent sex-typed behavior in boys and in girls. *Hormones and Behavior*, 61(4), 611–616.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
- Langlois, J. H., & Downs, A. C. (1980). Mothers, fathers, and peers as socialization agents of sex-typed play behaviors in young children. *Child Development*, 1237–1247.
- Lavelli, M., & Fogel, A. (2002). Developmental changes in mother-infant face-to-face communication: Birth to 3 months. *Developmental Psychology*, 38(2), 288.
- Leaper, C., Breed, L., Hoffman, L., & Perlman, C. A. (2002). Variations in the gender-stereotyped content of Children's television cartoons across genres. *Journal of Applied Social Psychology*, 32(8), 1653–1662.
- Leeb, R. T., & Rejskind, F. G. (2004). Here's looking at you, kid! A longitudinal study of perceived gender differences in mutual gaze behavior in young infants. *Sex Roles*, 50(1–2), 1–14.
- Li, R. Y. H., & Wong, W. I. (2016). Gender-typed play and social abilities in boys and girls: Are they related? *Sex Roles*, 74(9–10), 399. <https://doi.org/10.1007/s11199-016-0580-7>
- Lobel, T. E., & Menashri, J. (1993). Relations of conceptions of gender-role transgressions and gender constancy to gender-typed toy preferences. *Developmental Psychology*, 29(1), 150.
- Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin*, 109(2), 267.
- Martin, C. L., & Fabes, R. A. (2001). The stability and consequences of young children's same-sex peer interactions. *Developmental Psychology*, 37(3), 431–446. <https://doi.org/10.1037/0012-1649.37.3.431>
- Martin, C. L., & Ruble, D. (2004). Children's search for gender cues cognitive perspectives on gender development. *Current Directions in Psychological Science*, 13(2), 67–70.
- Martin, C. L., & Ruble, D. N. (2010). Patterns of gender development. *Annual Review of Psychology*, 61, 353.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126(3), 424.
- Meyer-Bahlburg, H. F. L., Dolezal, C., Baker, S. W., Carlson, A. D., Obeid, J. S., & New, M. I. (2004). Prenatal androgenization affects gender-related behavior but not gender identity in 5–12-year-old girls with congenital adrenal hyperplasia. *Archives of Sexual Behavior*, 33(2), 97–104. <https://doi.org/10.1023/B:ASEB.0000014324.25718.51>
- Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior & Development*, 6(2), 313–330.
- Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants a sex difference. *Psychological Science*, 19(11), 1063–1066.
- Nagy, E., Komppagne, H., Orvos, H., & Pal, A. (2007). Gender-related differences in neonatal imitation. *Infant and Child Development*, 16(3), 267–276.
- O'Brien, M., & Huston, A. C. (1985). Development of sex-typed play behavior in toddlers. *Developmental Psychology*, 21(5), 866.
- O'Brien, M., Huston, A. C., & Risley, T. R. (1983). Sex-typed play of toddlers in a day care center. *Journal of Applied Developmental Psychology*, 4(1), 1–9.
- Pasterski, V. L., Geffner, M. E., Brain, C., Hindmarsh, P., Brook, C., & Hines, M. (2005). Prenatal hormones and postnatal socialization by parents as determinants of male-typical toy play in girls with congenital adrenal hyperplasia. *Child Development*, 76(1), 264–278.

- Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, 19(11), 1067–11070.
- Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: Convergent evidence. *Infancy*, 19(1), 103–116.
- Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception-London*, 31(9), 1109–1122.
- Reeves, B. D. J., & Higgins, J. (2011). In: J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration, London, UK.
- Roopnarine, J. L. (1986). Mothers' and fathers' behaviors toward the toy play of their infant sons and daughters. *Sex Roles*, 14(1–2), 59–68.
- Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*, 44(4), 512.
- Schau, C. G., Kahn, L., Diepold, J. H., & Cherry, F. (1980). The relationships of parental expectations and preschool children's verbal sex typing to their sex-typed toy play behavior. *Child Development*, 266–270.
- Schwartz, S. H., & Rubel-Lifschitz, T. (2009). Cross-national variation in the size of sex differences in values: Effects of gender equality. *Journal of Personality and Social Psychology*, 97(1), 171.
- Schweinle, A., & Wilcox, T. (2004). Sex differences in infants' ability to represent complex event sequences. *Infancy*, 6(3), 333–359.
- Serbin, L. A., Connor, J. M., Burchardt, C. J., & Citron, C. C. (1979). Effects of peer presence on sex-typing of children's play behavior. *Journal of Experimental Child Psychology*, 27(2), 303–309.
- Serbin, L. A., Connor, J. M., & Citron, C. C. (1981). Sex-differentiated free play behavior: Effects of teacher modeling, location, and gender. *Developmental Psychology*, 17(5), 640–646.
- Servin, A., Bohlin, G., & Berlin, L. (1999). Sex differences in 1-, 3-, and 5-year-olds' toy-choice in a structured play-session. *Scandinavian Journal of Psychology*, 40(1), 43–48.
- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review*, 13(2), 147–183.
- Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 46(4), 849–856.
- Smith, P. K., & Daglish, L. (1977). Sex differences in parent and infant behavior in the home. *Child Development*, 1250–1254.
- Snow, M. E., Jacklin, C. N., & Maccoby, E. E. (1983). Sex-of-child differences in father-child interaction at one year of age. *Child Development*, 227–232.
- Stroup, D. F., Berlin, J. A., Morton, S. C., Olkin, I., Williamson, G. D., Rennie, D., ... Thacker, S. B. (2000). Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Jama*, 283(15), 2008–2012.
- The Cochrane Collaboration. (2011). Review Manager (RevMan), Version 5.1.., The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark, 2011.
- Todd, B. K., Barry, J. A., & Thommessen, S. A. O. (2016). Preferences for 'gender-typed' toys in boys and girls aged 9 to 32 months: Young children's 'gender-typed' toy preferences. *Infant and Child Development*. <https://doi.org/10.1002/icd.1986>
- Touwen, B. C. L. (1976). *Neurological development in infancy* (Vol. 58). London, UK: Heinemann Educational Books.
- Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development*, 14(4), 365–381.
- Turner, P. J., & Gervai, J. (1995). A multidimensional study of gender typing in preschool children and their parents: Personality, attitudes, preferences, behavior, and cultural differences. *Developmental Psychology*, 31(5), 759.
- Turner, P. J., Gervai, J., & Hinde, R. A. (1993). Gender-typing in young children: Preferences, behaviour, and cultural differences. *British Journal of Developmental Psychology*, 11, 323–342.
- United Nations Development Programme. Table 4: gender inequality index | Human development reports (n.d.). Retrieved from <http://hdr.undp.org/en/content/table-4-gender-inequality-index>
- Van de Beek, C., van Goozen, S. H., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2009). Prenatal sex hormones (maternal and amniotic fluid) and gender-related play behavior in 13-month-old infants. *Archives of Sexual Behavior*, 38(1), 6–15.
- Vygotsky, L. S. (1978). In M. Cole, et al. (Eds.), *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

- Wells, G. A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2011). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. www.ohri.ca/programs/clinical_epidemiology/oxford_web.ppt
- Wilansky-Traynor, P., & Lobel, T. E. (2008). Differential effects of an adult observer's presence on sex-typed play behavior: A comparison between gender-schematic and gender-aschematic preschool children. *Archives of Sexual Behavior*, 37(4), 548–557.
- Wilcox, T. (2003). Event-mapping tasks: Investigating the effects of prior information and event complexity on performance. *Infant Behavior & Development*, 26(4), 568–587.
- Wolf, T. M. (1973). Effects of live modeled sex-inappropriate play behavior in a naturalistic setting. *Developmental Psychology*, 9(1), 120.
- Yarrow, M. R., Campbell, J. D., & Burton, R. V. (1970). Recollections of childhood a study of the retrospective method. *Monographs of the Society for Research in Child Development*, iii–i83.
- Zosuls, K. M., Ruble, D. N., Tamis-LeMonda, C. S., Shrout, P. E., Bornstein, M. H., & Greulich, F. K. (2009). The acquisition of gender labels in infancy: Implications for gender-typed play. *Developmental Psychology*, 45(3), 688.
- Zosuls, K. M., Ruble, D. N., & Tamis-LeMonda, C. S. (2014). Self-socialization of gender in African American, Dominican immigrant, and Mexican immigrant toddlers. *Child Development*, 85(6), 2202–2217. <https://doi.org/10.1111/cdev.12261>

How to cite this article: Todd BK, Fischer RA, Di Costa S, et al. Sex differences in children's toy preferences: A systematic review, meta-regression, and meta-analysis. *Inf Child Dev*. 2017;e2064. <https://doi.org/10.1002/icd.2064>